Panel Data Models with Multiple Time-Varying Individual Effects

Seung C. Ahn∗
Arizona State University

Young H. Lee
Hansung University

Peter Schmidt
Michigan State University

This Version: October, 2006

Abstract

This paper considers a panel data model with time-varying individual effects. The data are assumed to contain a large number of cross-sectional units repeatedly observed over a fixed number of time periods. The model has a feature of the fixed-effects model in that the effects are assumed to be correlated with the regressors. The unobservable individual effects are assumed to have a factor structure. For consistent estimation of the model, it is important to estimate the true number of factors. We propose a generalized methods of moments procedure by which both the number of factors and the regression coefficients can be consistently estimated. Some important identification issues are also discussed. Our simulation results indicate that the proposed methods produce reliable estimates.

JEL Classification Codes: C51, D24
Keywords: panel data, time-varying individual effects, factor models.

Acknowledgment
The first author gratefully acknowledges the financial support of W.P. Carey School of Business and Dean's Council of 100 at Arizona State University, the Economic Club of Phoenix, and the alumni of W.P. Carey School. The third author acknowledges support under a Marie Curie Transfer of Knowledge Fellowship of the European Community under contract MTKD-CT-014288. We would like to thank seminar participants in Sungkunkwan University and Syracuse University.

∗ Corresponding Author: Seung C. Ahn, Department of Economics, W.P. Carey School of Business, Arizona State University, Tempe, AZ 85287; email: miniahn@asu.edu

1. Introduction

The use of panel data has been increasingly popular in empirical microeconomic and macroeconomic studies. An important advantage of using panel data is that researchers can obtain consistent or unbiased estimates of important parameters controlling for unobservable cross-sectional heterogeneity. An example of such heterogeneity, the so-called *individual effect*, is the effect of talent in a model of workers' hourly earnings. In order to estimate the effect of education on hourly wage rate consistently, researchers need to control for the heterogeneity in workers' talents or skills. Unfortunately, data containing information on individual workers' talents and skills are extremely rare. Without such information, it is extremely difficult, if not impossible, to control for talent using pure cross-sectional data. In contrast, when panel data are available, a variety of estimation methods (*e.g.*, Hausman and Taylor, 1981; Amemiya and MaCurdy, 1986; Cornwell and Rupert, 1988) can be used to control for the unobservable individual effects. Even if individual workers' talents are unobservable, it is possible to estimate the effect of education on hourly wage consistently.

In this paper we consider a more general panel data model in which the individual effect has multiple components and each of these components is time-varying. Specifically, the model assumes that the unobservable individual effects have a factor structure. For this model, we develop appropriate estimation and model-specification methods. Bai (2005) has considered the same panel factor model that we study in this paper (see also Bai and Ng, 2002; and Bai, 2003). His approach is designed for the analysis of panel data with large numbers of both time series and cross-section observations, and the regressors are assumed to be strictly exogenous to the random error terms in the model. Kneip, Sickles and Song (2005) also consider the same model but with the additional assumption that the factors change slowly and smoothly over time. Our paper is different from these papers in two respects. First, we focus on the case of panel data with a small number of time series observations and a large number of cross-section units (big N and small T). Second, we also consider the case in which some regressors are only weakly exogenous.

Standard panel data models assume that the unobservable individual effect is a single time-invariant component. However, this assumption may be excessively restrictive in practice. For example, consider a model of hourly wage rates. It is a well-known fact that labor productivity changes over the business cycle. Accordingly, the productivity of an individual's unobservable talent or skill would also change over the business cycle (Ahn, Lee and Schmidt, 2001). If so, the effect of unobservable talent on hourly wages would vary over time because workers' hourly wage rates depend on their labor productivity. It is also likely that hourly wage rates depend on multiple individual effects. For example, individual workers' wages would be affected by unexpected changes in macroeconomic variables due to changes in monetary or fiscal policies. However, the effects of these aggregate variables on wages would depend on individual-specific characteristics such as the worker's residential area and occupation. The panel data models that assume a single time-invariant individual effect are inappropriate for the analysis of data with such multiple time-varying individual effects.

There are many other examples of models that may require multiple time-varying effects. One example is the consumption model based on the life-cycle and rational-expectation hypothesis. This model predicts that current consumption growth depends on the unobservable marginal utility of expected life-time wealth. When consumers' future incomes are uncertain, their marginal utility of wealth varies over time (Altug and Miller, 1990; Pischke, 1995). Another example is the asset pricing models that assume time-varying risk premia (Campbell, 1987; Ferson and Foerester, 1994; Zhou, 1994). These models can be also viewed as panel data models with unobservable multiple time-varying individual effects. Finally, our approach can be used for the empirical studies of economic growth based on international data (*e.g.*, Mankiw, Romer and Weil, 1992; Islam, 1995; Caselli, Esquival and Lefort, 1996). Individual countries' economic growth rates could depend on world-wide supply shocks such as the oil shocks in the 1970's, and the technology shocks we have witnessed from the rapid development of the information technology industry in the 1990's. However, the effect of such world-wide shocks could depend on country-specific factors such as available human capital and natural resources.

The model we consider is also related to the issue of cross-sectional dependence, which is a growing research area. Many studies based on cross section data assume that the data are cross-sectionally independent. However, there are many cases in which the independence assumption is questionable. As we have discussed above, the decisions of individual economic agents (such as individuals, households, or firms) can depend on common macroeconomic shocks. When data contain such common factors, conventional estimators such as ordinary least squares (OLS) and instrumental-variables can be biased (Andrews, 2003). Even in the cases where such estimators are consistent, the estimated standard errors of the estimators obtained ignoring cross-sectional dependence could be seriously biased (for example, Chang, 2002). In response to these problems, many alternative estimation methods have been developed (Conley, 1999; Kelejian and Prucha, 1999; Chang, 2002). The method we develop in this paper provides an alternative solution for the analysis of panel data. Our model can allow cross-sectional dependence among individual effects.

Panel data models with time-varying individual effects and small numbers of time-series observations have been studied by Holtz-Eakin, Newey and Rosen (1988), Lee (1991), Chamberlain (1992), and Ahn, Lee and Schmidt (2001) [hereafter, ALS]. However, these studies, except Lee (1991), only consider cases with a single individual effect. Lee (1991) considered the case of multiple factors, but he made the unnecessarily strong assumption that the errors are i.i.d. normal, and he assumed that the true number of factors was known.

The goals of this paper are two-fold. The first is to investigate estimation methods that can produce consistent estimates under quite general assumptions. We accomplish this via GMM as opposed to nonlinear least squares. The second is to develop an estimation and testing procedure for the correct number of factors.

This paper is organized as follows. Section 2 introduces our model and assumptions. Section 3 develops the GMM estimation method for the model. Section 4 reports our simulation results. Some concluding remarks follow in section 5.

2. Basic Model

The basic model of this study is given by:

$$y_i = X_i\beta + u_i; \quad u_i = \eta_i + \varepsilon_i = \Xi\alpha_i + \varepsilon_i . \tag{1}$$

Here, i (= 1, ... , N) is the cross-sectional index and t (= 1, ... , T) is the time index. $y_i = (y_{i1},...,y_{iT})'$ is the $T\times 1$ vector of the dependent variables for individual i, $X_i = (x_{i1},...,x_{iT})'$ is the $T\times k$ matrix of time-varying regressors, β is the $k\times 1$ vector of coefficients to be estimated, and $\varepsilon_i = (\varepsilon_{i1},...,\varepsilon_{iT})'$ is the $T\times 1$ vector of random errors. Also η_i is the $T\times 1$ vector of individual effects and α_i is the $p\times 1$ vector of unobservable individual-specific variables. The $T\times p$ ($T > p$) matrix Ξ includes p time-varying factors that are common to all individuals. Note that the elements of η_i vary over time because the elements of Ξ vary over time. The unobservable individual effects have a "factor structure."

Where it helps the clarity of the argument, we will use subscript "o" to refer to "true value" of a parameter. For example, p_o refers to the true value of p. For the identification of the model (1), we need to define what we mean by the true number of factors, p_o. We assume that $rank(\Xi) = rank[E(\alpha_i\alpha_i')] = p_o$. That is, p_o is the smallest value of the number of factors that the factor representation $\Xi\alpha_i$ holds. The words "smallest value" are important because, if the factor representation $\Xi\alpha_i$ holds for a given p, it also holds for any greater number of factors than p. To see why, let Ξ^* be a $T\times(p+1)$ matrix of full column rank whose columns span the column space of Ξ. Then, there must exist a $(p+1)\times p$ matrix B such that $\Xi = \Xi^* B$. Thus, we have $\Xi\alpha_i = \Xi^*\alpha_i^*$, where $\alpha_i^* = B\alpha_i$ has (p+1) components. Observe that

$rank[E(\alpha_i^* \alpha_i^{*\prime})] \le p$. Thus, α_i^* contains some linearly redundant components. The assumption $rank[E(\alpha_i \alpha_i')] = p$ rules out such redundant components in η_i.

Notice that the rank assumption, $rank(\Xi) = rank[E(\alpha_i \alpha_i')] = p$, applies to $E(\alpha_i \alpha_i')$, not to $Var(\alpha_i)$. This is because we allow some components in α_i to be constant over *i*. As we discuss below, the model (1) cannot identify the effects of time-dummy variables unless we assume that $E(\alpha_i) = 0$. Thus, we need to estimate β leaving the time effects absorbed by $\Xi\alpha_i$. For such cases, one of the columns of Ξ may capture the time effects, the corresponding element in α_i being a constant.

In addition to the assumption $rank(\Xi) = p$, some normalization (parametric restrictions) on Ξ is necessary for identification because, for any nonsingular $p \times p$ matrix *C*, $\Xi\alpha_i = \Xi^* \alpha_i^*$, where $\Xi^* = \Xi C^{-1}$ and $\alpha_i^* = C\alpha_i$. This is the so-called "rotation" problem. A simple normalization we will impose on Ξ is that $\Xi = (\Theta', -I_p)'$, where Θ is a $(T\text{-}p) \times p$ matrix of unrestricted parameters. Imposing this normalization, we are reparameterizing the individual-effect components by $\Theta = -\Xi_1 \Xi_2^{-1}$, $\Xi^* = -\Xi\Xi_2^{-1}$ and $\alpha_i^* = -\Xi_2 \alpha_i$, where $\Xi = (\Xi_1', \Xi_2')'$.[1]

If the regressors x_{it} and the individual-specific effects α_i are uncorrelated, the coefficient vector β could be consistently estimated by ordinary least squares (OLS) or generalized least squares (GLS). We do not consider such trivial cases. Instead we consider

---

[1] The number of restrictions we impose on Ξ is p^2. In the factor analysis literature, the number of restrictions imposed on Ξ to avoid rotational indeterminacy is $p(p\text{-}1)/2$. The reason why we impose more restrictions on the model is that we do not impose any variance-covariance restrictions on the factors. The usual factor analysis assumes that the variance matrix of factors is an identity matrix. This assumption implies $p(p+1)/2$ restrictions. In total, the usual factor analysis imposes p^2 (= $p(p\text{-}1)/2 + p(p+1)/2$) restrictions on $Var(\Xi\alpha_i)$, which is the same number of the restrictions we use.

the cases in which X_i and α_i are correlated. Thus, the model (1) has the flavor of a fixed effects model. The relative sizes of N and T are important to determine an appropriate estimation method. Bai (2005) considers the estimation of the model (1) for the cases in which T is large and the regressors in X_i are stochastically independent of the errors. Instead, in this paper, we focus on the cases with relatively small T and large N, and thus asymptotics apply as $N \to \infty$ with T fixed.

To begin with, we assume that the model (1) does not contain time-dummy variables, or other variables that vary only over time, and are invariant over individuals. That is, we assume that x_{it} varies over different cross-section units. This is due to the identification issues that will be discussed below. We also do not include time-invariant individual characteristics as regressors, although the coefficients of such variables could be consistently estimated under certain circumstances. But we do use them as instruments in estimation. These variables will be the vector f_i below.

We define $z_i = (x_{i1}', x_{i2}', ..., x_{iT}', f_i')'$, our set of potential instrumental variables, and X = $(X_1', X_2', ..., X_N')'$. Using this notation, we make the following *Basic Assumptions* (BA):

(BA.1) $(z_i', \alpha_i', \varepsilon_i')'$ is independent and identically distributed over i, with finite moments up to fourth order.

(BA.2) $E(z_i z_i')$ is nonsingular.

(BA.3) $rank[E(z_i \alpha_i')] = rank[E(\alpha_i \alpha_i')] = p_o$.

(BA.4) $E(\varepsilon_i \mid f_i, \alpha_i) = 0_{T\times 1}$.

(BA.5) $(X, I_N \otimes \Xi_o)$ is of full column rank.

Several comments on BA follow. (BA.1) is a technical assumption for the consistency and asymptotic normality of the estimators discussed below. (BA.1) assumes that data are cross-sectionally independent. However, we can relax this assumption for z_i and α_i. What we really need are just conditions under which the law of large numbers holds for $(z_i', \alpha_i')'$; that is, $p\lim_{N\to\infty} N^{-1}\Sigma_{i=1}^{N}(z_i', \alpha_i')'(z_i', \alpha_i') = E[(z_i', \alpha_i')'(z_i', \alpha_i')]$.[2]

(BA.2) implies that there is no perfect multicollinearity in z_i.

(BA.3) is intended to capture the essential feature of a fixed effects model, in the sense that α_i is allowed to be correlated with z_i in an unrestricted way. However, it says more than that since it implies α_i *must* be correlated with z_i. (BA.3) is an important identification condition for the parameter matrix Ξ (or Θ). If all of the variables in z_i are uncorrelated with both α_i and ε_i, there is no way to discriminate α_i from ε_i unless some strong restrictions are imposed on the variances and covariances of the components in α_i and ε_i (such as the diagonality of $Var(\varepsilon_i)$).[3] Notice that (BA.3) applies to $E(z_i\alpha_i')$, not to the covariances between z_i and α_i. Thus, we allow cases in which some components of α_i are constant over time.

(BA.4) assumes that all of the variables in f_i and α_i are uncorrelated with any error terms in $\varepsilon_i = (\varepsilon_{i1}, ..., \varepsilon_{iT})'$. But the errors ε_{it} are allowed to be autocorrelated or heteroskedastic over time. Assumptions about the correlations between ε_i and X_i will be made below.

Finally, Assumption (BA.5) rules out certain types of exact multicollinearity that will be discussed in detail below.

For consistent estimation of the model (1), we need to make assumptions about the

---

[2] We may also allow the error vectors ε_i to be cross-sectionally correlated as in Bai (2005). While here we do not consider cases with cross-sectionally dependent data, it could be shown that the GMM estimators we discuss below are still consistent and asymptotically normal. Also the asymptotic variance matrices of the estimators could be consistently estimated by the method of Conley (1999).

[3] In general, we cannot identify the columns of Ξ that correspond to the components in α_i uncorrelated with regressors. However, this is not a serious problem in estimating β. What we need to control for to obtain a consistent estimator of β are the components in α_i that are correlated with regressors. We can treat the components in α_i uncorrelated with regressors (and the corresponding columns of Ξ) as parts of the random error vector ε_i.

relationship between the time-varying regressors X_i and the random noise error terms ε_{it}. We consider two types of regressors: strictly exogenous (SE) and weakly exogenous (WE) regressors. In order to distinguish between these regressors, we partition x_{it} into $x_{it} = (x'_{S,it}, x'_{W,it})'$, where

(SE) $E(\varepsilon_{it} \mid x_{S,i1}, x_{S,i2}, ..., x_{S,iT}) = 0$, $t = 1, \ldots, T$.

(WE) $E(\varepsilon_{it} \mid x_{W,i1}, ..., x_{W,it}) = 0$, $t = 1, \ldots, T$.

With these assumptions made, we can be more explicit about the relationship between our model and previous models. If we view α_i and Θ as multiple factors and a factor loading matrix, respectively, we can easily see that the model (1) is similar to the multi-factor model frequently used in empirical finance (Campbell, Lo and Mackinlay, 1997, Chapter 6). However, the model (1) is different from the multi-factor model in two aspects. First, while the multi-factor model treats the factors (α_i) as random variables uncorrelated with the regressors, the model (1) allows non-zero correlations between regressors and factors. This treatment is essentially the same as the assumption that the α_i are unobservable parameters (Mundlak, 1978) and therefore corresponds to a fixed-effects as opposed to random-effects treatment. Second, we do not assume a diagonal variance matrix for the error terms.

Lee (1991) and Lee and Schmidt (1993) proposed estimation of the model by MLE, assuming that the errors ε_{it} are i.i.d. normal. This leads to a nonlinear least squares estimator. However, when T is small (fixed), this is a non-regular problem. As ALS showed, the consistency of the nonlinear least squares estimator depends on the errors being white noise (normality is not required), and even if the errors are white noise the nonlinear least squares estimator is inefficient, and the variance matrix of the estimate has the sandwich-form because the usual information equality does not hold. When both N and T are large (asymptotics are as N and T both go to infinity), Bai (2003, 2005) shows that nonlinear least squares is consistent without the white noise assumption. Bai (2005) also shows that the methods of Bai and Ng

(2002) can be used to consistently estimate the true number of factors.

This paper follows ALS by using a GMM approach that does not rely on the white noise assumption and that is consistent when T is fixed. This paper differs from ALS in that we now consider multiple factors and correspondingly we propose methods for choosing the number of factors consistently. This paper also differs from ALS, and from Bai (2005), in that we allow some regressors to be weakly exogenous. Thus our approach can be applied to the dynamic panel data model or other models with weakly exogenous regressors.

3. Estimation and Tests

In this section, we consider the GMM estimation of the model (1). We also examine two methods that can be used to estimate the true number of individual effects α_i (p_o).

We begin with the estimation of β and Θ when p_o is known. For given p, we define

$$H(\theta) = [H_1(\theta_1), H_2(\theta_2), ..., H_{T-p}(\theta_{T-p})] = (I_{T-p}, \Theta)', \tag{2}$$

where θ_j (j = 1, ... , T-p) is the j'th column of Θ', $\theta = vec(\Theta')$, and $H_j(\theta_j)$ is the j'th column of $H(\theta)$. Observe that the first (j-1) entries of $H_j(\theta_j)$ equal zero. This property of the columns of $H(\theta)$ will be important to understand the moment conditions we can utilize for the cases with weakly exogenous regressors.

Note that $H(\theta)'\Xi = (I_{T-p}, \Theta)(\Theta', -I_T)' = 0_{(T-p)\times p}$. Thus, pre-multiplying the model (1) by $H(\theta_o)$, we can remove the unobservable individual effects from the model:

$$H(\theta_o)' y_i = H(\theta_o)' X_i \beta_o + H(\theta_o)' \varepsilon_i, \tag{3}$$

where,

$$H(\theta)'\varepsilon_i = \begin{pmatrix} H_1(\theta_1)'\varepsilon_i \\ H_2(\theta_2)'\varepsilon_i \\ \vdots \\ H_{T-p}(\theta_{T-p})'\varepsilon_i \end{pmatrix} = \begin{pmatrix} \varepsilon_{i1} + \theta_1'\varepsilon_{i,*} \\ \varepsilon_{i2} + \theta_2'\varepsilon_{i,*} \\ \vdots \\ \varepsilon_{i,T-p} + \theta_{T-p}'\varepsilon_{i,*} \end{pmatrix}; \ \varepsilon_{i,*} = \begin{pmatrix} \varepsilon_{i,T-p+1} \\ \varepsilon_{i,T-p+2} \\ \vdots \\ \varepsilon_{iT} \end{pmatrix}.$$

In the following two subsections, we consider the GMM estimation of the model (3) with strictly and/or weakly exogenous regressors. Some identification issues will be discussed in detail in the third subsection.

3.1. Estimation and Testing Using Strictly Exogenous Instruments

We define:

$$w_{S,i} = (x'_{S,i1},\ldots,x'_{S,iT},f'_i)'. \tag{4}$$

Here the subscript "S" indicates that the variables in x_S satisfy the strict exogeneity condition (SE) above. Clearly, under BA and SE, all of the variables in $w_{S,i}$ are uncorrelated with all of the errors in ε_i. That is,

$$E(\varepsilon_i \otimes w_{S,i}) = 0_{Tq\times 1},$$

where q is the number of the variables in $w_{S,i}$. At $p = p_o$, these instrumental variables satisfy the following moment conditions:

$$E[m_{S,i}(\xi_o)] \equiv E[H(\theta_o)'u_i(\beta_o) \otimes w_{S,i}] = 0_{(T-p)q\times 1}, \tag{5}$$

where $\xi = (\beta',\theta')'$, and $u_i(\beta) = y_i - X_i\beta$. Thus, the parameter vector ξ can be estimated by GMM applied to the moment conditions in (5).

Let p denote the number of factors chosen for estimation. The consistent estimation of p_o, θ_o and β_o requires the following three identification conditions (I) to hold:

(I.1) When $p < p_o$, there exist no values of θ and β that satisfy (5).

(I.2) When $p = p_o$, $\theta = \theta_o$ and $\beta = \beta_o$ are the unique solutions for (5).

(I.3) When $p > p_o$, there exist some θ and β that satisfy (5).

Assumptions (BA.3) and (BA.5) imply that all of these three conditions hold. With $p < p_o$, any

H matrix cannot satisfy (5), because there is no T×(T-p) matrix H that can completely remove $\Xi_o\alpha_i$ from $u_i(\beta_o)$. Thus, under BA, for any H with $p < p_o$,

$$\begin{aligned} E[w_{S,i}u_i(\beta)'H] &= E[w_{S,i}u_i(\beta_o)'H + w_{S,i}(\beta_o - \beta)'X_i'H] \\ &= E(w_{S,i}\alpha_i')\Xi_o'H + E[w_{S,i}(\beta_o - \beta)'X_i'H] \neq 0, \end{aligned}$$

even if $\beta = \beta_o$. Thus, (I.1) holds as long as $E(w_{S,i}\alpha_i')$ is of full column.

Assumption (BA.5), as well as (BA.3), is crucial for (I.2) to hold. To see why, we here consider two cases in which the assumption is violated, although more detailed discussions of these cases will be given in subsection 3.3. The first and obvious case is the case in which the model (1) contains time-invariant regressors *and* Ξ_o contains a time-invariant component.[4] Suppose that $y_i = X_i\beta_o + e_T f_i'\gamma_o + (\Xi_o\alpha_i + \varepsilon_i)$, where e_T is a T×1 vector of ones and γ_o is the vector of the coefficients on the time-invariant regressors f_i. For this case, $H_o = H(\theta_o)$ is orthogonal to e_T. Then, we easily see that the conditions (5) hold for any value of γ. Therefore the parameter vector γ is not identifiable. This is the same problem as in the usual fixed-effects model in which the coefficients of time-invariant regressors are not identified. Of course, γ can be identified if we can assert on *a priori* grounds that there is no time-invariant component in Ξ. However, if such *a priori* information is not available, and if the effects of time-invariant regressors are not of research interest, it may be advisable to estimate the model (1) without using time-invariant variables as regressors so that the individual effects $\Xi\alpha_i$ absorb the effects of the time-invariant regressors (and thus have a time-invariant component). Some further treatment of cases with both time-invariant regressors and time-invariant individual effects is given in subsection 3.3.

The second case of lack of identification is the case in which X_i contains time-dummy variables. Suppose that $y_i = \lambda_o + X_i\beta_o + (\Xi_o\alpha_i + \varepsilon_i)$, where λ is the T×1 vector of the time

---

[4] The meaning of this phrase merits some explanation. The most obvious interpretation would be that a column of Ξ is a vector of ones. Given our normalization, this cannot be true. The proper meaning is that some linear combination of columns of Ξ_o is equal to one for all t. This is equivalent to saying that H_o is orthogonal to e_T.

effects. Let $\lambda^* = \lambda_o + \Xi_o d$, so that $H_o'\lambda^* = H_o'\lambda_o$ for any $p\times 1$ vector d. Then, it is clear that whenever λ_o satisfies (5), so does λ^*. This implies that the GMM estimator based on (5) cannot identify the time effects. The non-identification of the effects of the time-dummy variables does not cause any problem in estimating the regression coefficient vector β. We can simply estimate β leaving the time effects to be captured by the factors in Ξ.

Notice that (I.3) allows multiple solutions for (5). Indeed, for any $p > p_o$ there are infinitely many θ's that satisfy (5). For example, consider a case with $T = 3$ and $p_o = 1$. For this case, we have

$$\Xi_o = (\theta_{1,o}, \theta_{2,o}, -1)'; \; H(\theta_o)' = \begin{pmatrix} 1 & 0 & \theta_{1,o} \\ 0 & 1 & \theta_{2,o} \end{pmatrix}.$$

Suppose now that for estimation, we choose $p = 2$ ($> p_o$) and $H^*(\theta^*) = (1, \theta_1^*, \theta_2^*)'$. Then, it can be shown that $H^*(\theta^*)$ satisfies (5) for any θ_1^* if we set $\theta_2^* = \theta_{1,o} + \theta_{2,o}\theta_1^*$.

The usual identification condition for GMM is that the moment conditions used should hold only at the true values of parameters. But due to (I.3), our identification conditions are weaker than this usual identification condition. When $p > p_o$ is chosen for GMM, we are unable to consistently estimate θ. Thus, if θ, as well as β, is a parameter vector of research interest, it is important to use the true value of p for GMM based on the moment conditions (5).[5] Below we introduce two methods by which p_o can be consistently estimated.

When $p = p_o$ is chosen, the identification of θ_o and β_o requires that the number of the moment conditions in (5) should be no less than the number of parameters in ξ:

$$(T-p)q - [(T-p)p + k] = (T-p)(q-p) - k \geq 0.$$

[5] Ahn, Lee and Schmidt (2005) use this model to estimate the time-varying technical inefficiencies of Indonesian rice farms. In the paper, $\Xi\alpha_i$ measures farm i's technical inefficiency. Thus, the consistent estimation of farms' inefficiencies requires a consistent estimate of Ξ (or θ).

For this condition to hold, $w_{S,i}$ must be chosen such that $q > p$. However, if q is large, imposing all of the moment conditions (5) in GMM might not be desirable, because it might cause the finite-sample properties of the GMM estimator to deteriorate. In practice, it may be more desirable to use only a subset of the instruments $w_{S,i}$. In the next section, we will investigate the finite-sample properties of the GMM estimators utilizing all and some of the instruments in $w_{S,i}$.

Under (5) and the assumption that the value of p is correctly chosen, the optimal GMM estimator of ξ, $\tilde{\xi}_S$, among those utilizing the moment conditions in (5), can be obtained by minimizing:

$$J_{S,N}(\xi|p) \equiv N\left(\frac{1}{N}\Sigma_{i=1}^{N} m_{S,i}(\xi)\right)'\left(\frac{1}{N}\Sigma_{i=1}^{N} m_{S,i}(\hat{\xi})m_{S,i}(\hat{\xi})'\right)^{-1}\left(\frac{1}{N}\Sigma_{i=1}^{N} m_{S,i}(\xi)\right), \qquad (6)$$

where $\hat{\xi}$ is an initial consistent estimator of ξ. Observe that the weighting matrix $[N^{-1}\Sigma_{i=1}^{N} m_{S,i}(\hat{\xi})m_{S,i}(\hat{\xi})']^{-1}$ is a consistent estimator of the inverse of $N^{-1}\Sigma_{i=1}^{N} Var[m_{S,i}(\xi_o)]$. Using this optimal GMM estimator, we can also test for the true value of p (p_o). The overidentifying restriction test statistic by Hansen (1982) is given by $J_{S,N}(\tilde{\xi}_S|p)$. This statistic has the following properties:

Proposition 1: Suppose that the model (1) satisfies BA. Then, as $N \to \infty$,

(i) for $p = p_o$, $J_{S,N}(\tilde{\xi}_S|p) \to_d \chi^2[(T-p_o)(q-p_o)-k]$, and

(ii) for $p < p_o$, $J_{S,N}(\tilde{\xi}_S|p) \to_p \infty$,

where "$\to_d$" means "converges in distribution", and "$\to_p$" means "converges in probability".

The proof of Proposition 1 is omitted. The result (i) is straightforward from Hansen

(1982). The result (ii) is obtained because $p\lim_{N\to\infty} J_{S,N}(\tilde{\xi}_S|p)/N$ is a finite positive number. Note that Proposition 1 explains the asymptotic distribution of $J_{S,N}(\tilde{\xi}_S|p_o)$ only for the cases of $p \le p_o$. When $p > p_o$, there are infinitely many θ's that satisfy (5) as discussed above. Therefore, the estimator $\tilde{\xi}_S$ suffers from an identification problem. For this case, the distribution of the statistic is generally unknown.

Based on Proposition 1, we can consistently estimate p_o using a sequential testing method. We may begin by testing the hypothesis of $p_o = 0$ against the alternative hypothesis of $p_o > 0$. If the null hypothesis is rejected, we move to test the null hypothesis of $p_o = 1$ against the alternative of $p_o > 1$. We continue this procedure until the null hypothesis is not rejected. Let $\tilde{p}$ be the smallest p that we do not reject by the $J_{S,N}(\tilde{\xi}_{S,p}|p)$ statistic. For the consistency of $\tilde{p}$, the significance level used for the procedure should be appropriately adjusted (Bauer, Pötscher and Hackl, 1988; and Cragg and Donald, 1997). Following the proof of Theorem 5 of Cragg and Donald (1997), we can easily obtain the following result:

Proposition 2: Let b_N be the significance level used for the sequential test method. Choose b_N such that $b_N \to 0$ and $-\ln(b_N)/N \to 0$ as $N \to \infty$. Then, $\tilde{p}$ is consistent.

Bai and Ng (2002) and Bai (2005) develop alternative methods to estimate p for the model with both large N and T. Cragg and Donald (1997) use similar methods to estimate ranks of matrices. We can also estimate p using their methods. The criterion functions we can use to estimate p are of the form:

$$S_N(p) = J_{S,N}(\tilde{\xi}_S|p) - f(N)g(p). \tag{7}$$

With appropriate choices of $f(N)$ and $g(p)$, a consistent estimate of p can be obtained by minimizing the criterion function $S_N(p)$. There are many possible choices of $f(N)$ and $g(p)$.

Among them, we consider the following Schwarz Criterion:

Schwarz Criterion (BIC): $f(N) = ln(N)$ and $g(p) = a\{(T-p)(q-p)-k\}$.

Here, a is an arbitrarily chosen positive number. Observe that under BIC, (i) $f(N) \to \infty$, and (ii) $f(N)/N \to \infty$, as $N \to \infty$. The function $g(p)$ is simply the degrees of freedom of $J_{S,N}(\tilde{\xi}_S|p)$. Let $\widehat{p}$ be the value of p which minimizes the function (7) with the BIC choices of $f(N)$ and $g(p)$. The condition (i) is required to warrant that $p\lim_{N\to\infty} \Pr(\widehat{p} > p_o) = 0$, while the condition (ii) is for $p\lim_{N\to\infty} \Pr(\widehat{p} < p_o) = 0$ (Cragg and Donald, 1997). Observe that for $\widehat{p}$ to be greater than p_o, $S_N(p_o) - S_N(p^*) > 0$ for some $p^* > p_o$. Thus, $\Pr(\widehat{p} > p_o) \leq \Pr[S_N(p_o) - S_N(p^*) > 0]$. Note also that

$$\begin{aligned}\Pr[S_N(p_o)-S_N(p^*) > 0] &= \Pr[J_{S,N}(\tilde{\xi}_S \mid p_o)-J_{S,N}(\tilde{\xi}_S \mid p^*)+f(N)(g(p^*)-g(p_o)) > 0] \\ &\leq \Pr[J_{S,N}(\tilde{\xi}_S \mid p_o)+f(N)(g(p^*)-g(p_o)) > 0] \to 0,\end{aligned}$$

since $J_{S,N}(\tilde{\xi}_S \mid p_o)$ is a χ^2 random variable that is bounded almost surely, $(g(p^*)-g(p_o))$ is a fixed negative number, and $f(N) \to \infty$. Thus, $\Pr(\widehat{p} > p_o) \to 0$.[6] Similarly, for any $p^* < p_o$,

$$\begin{aligned}\Pr[S_N(p_o)-S_N(p^*)>0] &= \Pr\left[\frac{1}{N}(J_{S,N}(\tilde{\xi}_S|p_o)-J_{S,N}(\tilde{\xi}_S|p^*))+\frac{f(N)}{N}(g(p^*)-g(p_o))>0\right] \\ &\to 0,\end{aligned}$$

because $f(N)/N \to 0$, $J_{S,N}(\tilde{\xi}_S|p_o)/N \to_p 0$, and $J_{S,N}(\tilde{\xi}_S|p^*)/N$ converges in probability to a fixed positive number in probability. This implies that $\Pr(\widehat{p} < p_o) \to 0$. Thus, we obtain the following result.

Proposition 3: Suppose the model (1) satisfies BA. Then $\widehat{p}$ is consistent.

---

[6] Notice that this proof does not required the uniqueness of θ and β when $p > p_o$. Proposition 2 still holds even if θ and β are not unique for $p > p_o$.

One remark follows on Propositions 2 and 3. When the optimal weighting matrix $[N^{-1}\Sigma_{i=1}^{N} m_{S,i}(\hat{\xi}) m_{S,i}(\hat{\xi})']^{-1}$ is used for the GMM estimation of ξ as in (6), both the sequential test and the criterion-function methods described in the two propositions respectively can be used to consistently estimate p_o. However, the former method could not be used when the parameter vector ξ is not estimated with the optimal weighting matrix. In contrast, the latter method does not require use of the optimal weighting matrix. Notice that Proposition 3 holds as long as $J_{S,N}(\tilde{\xi}_S \mid p_o)$ is an almost surely bounded random variable. Indeed, the statistic $J_{S,N}(\tilde{\xi}_S \mid p_o)$ computed with a non-optimal weighting matrix is asymptotically a weighted average of independent chi-squared random variables, which is clearly bounded (see, for example, Jagannathan and Wang, 1996).

The computation of $\tilde{\xi}_S$ and $J_{S,N}(\tilde{\xi}_S|p)$ requires an initial consistent estimator of ξ that can be used to estimate the optimal weighting matrix. We now consider a simple iterative procedure by which a consistent estimator of ξ can be easily obtained. Under some circumstances, this iterative procedure can also lead to optimal GMM estimators which are asymptotically equivalent to $\tilde{\xi}_S$.

A consistent GMM estimator can be obtained by minimizing (6) replacing the weighting matrix by any conformable and asymptotically nonstochastic positive definite matrix. Consider the GMM estimator which minimizes the following function:

$$r_{S,N}(\xi|A,p)=N\left(\frac{1}{N}\Sigma_{i=1}^{N}\left(H'u_i(\beta)\otimes w_{S,i}\right)\right)'\left(H'AH\otimes\hat{\Omega}_{ww,S}\right)^{-1}\left(\frac{1}{N}\Sigma_{i=1}^{N}\left(H'u_i(\beta)\otimes w_{S,i}\right)\right), \quad (8)$$

where $H = H(\theta)$, $\hat{\Omega}_{ww,S} = N^{-1}\Sigma_{i=1}^{N} w_{S,i} w_{S,i}'$, and A is a $T\times T$ positive definite matrix. Observe that the weighting matrix $\left(H'AH \otimes \hat{\Omega}_{ww,S}\right)^{-1}$ contains the unknown parameter vector θ. The usual two-step and iterative GMM procedures estimate the parameters in the weighting matrix and the moment function (for example, $\left(H'AH \otimes \hat{\Omega}_{ww,S}\right)^{-1}$ and $N^{-1}\Sigma_{i=1}^{N} m_{S,i}(\xi)$, respectively, in our case) sequentially. In contrast, we estimate θ in the moment functions in $m_{S,i}(\xi)$ and the weighting matrix $\left(H'AH \otimes \hat{\Omega}_{ww,S}\right)^{-1}$ concurrently. Thus, the GMM estimator minimizing (8) has a flavor of the continuous-updating estimator of Hansen, Heaton and Yaron (1996).

The relationship between continuous-updating GMM and the usual two-step GMM in the estimation of model (1) is the same as the relationship between Limited Information Maximum Likelihood (LIML) and Two-Stage Least Square (2SLS) in the estimation of a system of simultaneous equations. An important advantage of the LIML estimator over 2SLS is that it is invariant to the normalization used to express the simultaneous equations. The continuous-updating GMM estimator that minimizes (8) has the similar property: It is invariant to the normalization of Ξ.

While the function (8) is highly nonlinear, the estimation procedure can be simplified as follows. Define:

$$\hat{\Omega}_{wu,S}(\beta) = \frac{1}{N}\Sigma_{i=1}^{N} w_{S,i}(y_i - X_i\beta)'.$$

Then, it can be shown that the criterion function (8) equals:

$$N \times trace\left(H'\hat{\Omega}_{wu,S}(\beta)'(\hat{\Omega}_{ww,S})^{-1}\hat{\Omega}_{wu,S}(\beta)H(H'AH)^{-1}\right). \tag{9}$$

While the matrix H is a function of θ, we can consider the minimization of (9) with respect to H and β. For given β, the H matrix that minimizes (9) is the matrix of the eigenvectors corresponding to the (T-p) smallest eigenvalues of the matrix

$$A^{-1}\hat{\Omega}_{wu,S}(\beta)'(\hat{\Omega}_{ww,S})^{-1}\hat{\Omega}_{wu,S}(\beta).$$

(See Magnus and Neudecker, 1999, Chapter 11). Given H, the β value minimizing (8) or (9) is obtained by:

$$\left(\left(\frac{1}{N}\Sigma_{i=1}^{N}(H'X_i \otimes w_{S,i})\right)'\left(H'AH \otimes \hat{\Omega}_{ww,S}\right)^{-1}\left(\frac{1}{N}\Sigma_{i=1}^{N}(H'X_i \otimes w_{S,i})\right)\right) \tag{10}$$
$$\times\left(\frac{1}{N}\Sigma_{i=1}^{N}(H'X_i \otimes w_{S,i})\right)'\left(H'AH \otimes \hat{\Omega}_{ww,S}\right)^{-1}\left(\frac{1}{N}\Sigma_{i=1}^{N}(H'y_i \otimes w_{S,i})\right).$$

Thus, we can obtain consistent estimates of β and H by computing β and H iteratively until the results converge. Let $\hat{H}$ be the estimate of the unrestricted H obtained from this iterative procedure. Partition $\hat{H} = (\hat{H}_{T-p}{}', \hat{H}_p{}')'$ such that $\hat{H}_{T-p}$ is a (T-p)×(T-p) square matrix. Then, $\hat{\Theta}' = \hat{H}_p(\hat{H}_{T-p})^{-1}$ is the value of Θ' that minimizes (8).

The GMM estimator obtained by the above iterative procedure is suboptimal. This is so because the weighting matrix used in (8) is suboptimal. However, an optimal GMM estimator can be obtained by the above iterative procedure if the instruments in $w_{S,i}$ satisfy the following assumption, which we call the "no conditional heteroskedasticity" (NCH) condition:

(NCH) $\quad E(\varepsilon_i\varepsilon_i{}' \mid w_{S,i}) = \Sigma_\varepsilon$, for all i.

This condition holds if the error terms in ε_i are stochastically independent of the variables in $w_{S,i}$. Under this assumption,

$$Var\left(H_o'u(\beta_o) \otimes w_{S,i}\right) = H_o'\Sigma_\varepsilon H_o \otimes E(w_{S,i}w_{S,i}') = H_o'\Sigma H_o \otimes E(w_{S,i}w_{S,i}'), \tag{11}$$

where $H_o = H(\theta_o)$ and $\Sigma = E[u_i(\beta_o)u_i(\beta_o)']$. Thus, the optimal weighting matrix will have the form of $[H'\hat{\Sigma}H \otimes \hat{\Omega}_{ww,S}]^{-1}$, where $\hat{\Sigma} = N^{-1}\Sigma_{i=1}^{N}u_i(\hat{\beta})u_i(\hat{\beta})'$ is a consistent estimator of Σ. An optimal GMM estimator of ξ, say $\ddot{\xi}_S = (\ddot{\beta}_S{}', \ddot{\theta}_S{}')'$, can be obtained by minimizing (8) with the matrix A replaced by $\hat{\Sigma}$. In addition, $r_{S,N}(\ddot{\xi}_S|\hat{\Sigma},p)$ can be used as a model specification test statistic. This statistic simply equals the sum of the first (T-p) smallest eigenvalues of the matrix

$$(\hat{\Sigma})^{-1}\hat{\Omega}_{wu,S}(\ddot{\beta}_S)'(\hat{\Omega}_{ww,S})^{-1}\hat{\Omega}_{wu,S}(\ddot{\beta}_S),$$

and is asymptotically equivalent to $J_{S,N}(\tilde{\xi}_S|p)$ when NCH holds. An empirical application of this iterative method appears in a companion paper of ours (Ahn, Lee and Schmidt, 2005).

3.2 Estimation and Testing Using Weakly Exogenous Instruments

When weakly exogenous instruments are available (in addition to strictly exogenous ones), a more (asymptotically) efficient estimator can be obtained by using them in GMM. Define:

$$w_{j,i} = (w'_{S,i}, x'_{W,i1}, \ldots, x'_{W,ij})', \; j = 1, \ldots, T\text{-}p. \tag{12}$$

where $w_{j,i}$ contains q_j variables. Observe that all of the variables in $w_{j,i}$ are also included in $w_{j+1,i}$. However, in practice, one does not need to keep this structure of the instruments. Only subsets of $w_{j,i}$ could be used for GMM.

Clearly, under BA, SE and WE,

$$E(\varepsilon_{is} \mid w_{j,i}) = 0, \; s \geq j, \; j = 1, \ldots, T\text{-}p. \tag{13}$$

Thus, we have the following moment conditions at $p = p_o$:

$$E(m_{W,i}(\xi_o)) \equiv E\begin{pmatrix} H_1(\theta_{1,o})' u_i(\beta_o) w_{1,i} \\ H_2(\theta_{2,o})' u_i(\beta_o) w_{2,i} \\ \vdots \\ H_{T-p}(\theta_{T-p,o})' u_i(\beta_o) w_{T-p,i} \end{pmatrix} = 0. \tag{14}$$

These moment conditions hold because $H_j(\theta_{j,o})' u_i(\beta_o)$ does not depend on $\varepsilon_{i1}, \ldots, \varepsilon_{i,j-1}$ (by the structure of H given in (2)), so that $E(H_j(\theta_o)' u_i(\beta_o) w_{j,i}) = 0$.

Differently from cases with only strictly exogenous instruments, a simple continuous-updating GMM procedure is not immediately available. Instead, we here consider a two-step GMM procedure that can also lead to an optimal estimator under some conditions. For this procedure, let $D = (d_1, \ldots, d_{T-p})$ be an arbitrary $(T\text{-}p)\times(T\text{-}p)$ invertible lower triangular matrix. Let $\Theta' D = \Theta^{D\prime}$ and $\theta^D = vec(\Theta^{D\prime})$. With this notation, we can define:

$$H^D(\theta^D) = (H_1^D(\theta_1^D),...,H_{T-p}^D(\theta_{T-p}^D)) \equiv (H(\theta)d_1,...,H(\theta)d_{T-p}) = H(\theta)D\,.$$

Observe that by the construction of $H^D(\theta^D)$, $H_j^D(\theta_j^D)' u_i(\beta_o)$ remains uncorrelated with the instruments in $w_{j,i}$. Thus, at $p = p_o$, we have the following moment condition:

$$E(m_{W,i}^D(\xi_o^D)) \equiv E\begin{pmatrix} H_1^D(\theta_{1,o}^D)' u_i(\beta_o) w_{1,i} \\ H_2^D(\theta_{2,o}^D)' u_i(\beta_o) w_{2,i} \\ \vdots \\ H_{T-p}^D(\theta_{T-p,o}^D)' u_i(\beta_o) w_{T-p,i} \end{pmatrix} = 0\,, \tag{15}$$

where $\xi^D = (\beta', \theta^{D\prime})'$. In general, the GMM estimator based on the moment condition (14) is different from the estimator based on (15). However, if all of the variables in $w_{j,i}$ are included in $w_{j+1,i}$, the conditions (15) are one-to-one transformations of those in (14). Thus, for such cases, the optimal GMM estimator based on (14) should be asymptotically equivalent to its counterpart based on (15).

A consistent GMM estimator based on the moment conditions (15) can be obtained by minimizing the following criterion function:

$$\begin{aligned} r_{W,N}(\xi^D|p) &= N\left(\frac{1}{N}\Sigma_{i=1}^N m_{W,i}^D(\xi^D)\right)'\left(diag(\hat{\Omega}_{ww,1},...,\hat{\Omega}_{ww,T-p})\right)^{-1}\left(\frac{1}{N}\Sigma_{i=1}^N m_{W,i}^D(\xi^D)\right) \\ &= N\Sigma_{j=1}^{T-p} H_j^D(\theta_j^D)'\hat{\Omega}_{wu,j}(\beta)'\left(\hat{\Omega}_{ww,j}\right)^{-1}\hat{\Omega}_{wu,j}(\beta) H_j^D(\theta_j^D), \end{aligned} \tag{16}$$

where $\hat{\Omega}_{ww,j} = N^{-1}\Sigma_{i=1}^N w_{j,i} w_{j,i}'$ and $\hat{\Omega}_{wu,j}(\beta) = N^{-1}\Sigma_{j=1}^N w_{j,i} u_i(\beta)'$. This function can be minimized by an iterative procedure. For given $H^D(\theta^D)$, the value of β minimizing $r_{W,N}(\xi^D|p)$ equals

$$\left[\Sigma_{j=1}^{T-p}\left(\frac{1}{N}\Sigma_{i=1}^{N}X_i'H_j^D(\theta_j^D)w_{j,i}'\right)\left(\hat{\Omega}_{ww,j}\right)^{-1}\left(\frac{1}{N}\Sigma_{i=1}^{N}w_{j,i}H_j^D(\theta_j^D)'X_i\right)\right]^{-1}$$
$$\times\Sigma_{j=1}^{T-p}\left(\frac{1}{N}\Sigma_{i=1}^{N}X_i'H_j^D(\theta_j^D)w_{j,i}'\right)\left(\hat{\Omega}_{ww,j}\right)^{-1}\left(\frac{1}{N}\Sigma_{i=1}^{N}w_{j,i}H_j^*(\theta_j^*)'y_i\right).$$

Define $\hat{\Delta}_j(\beta)=\hat{\Omega}_{wu,j}(\beta)'(\hat{\Omega}_{ww,j})^{-1}\hat{\Omega}_{wu,j}(\beta)$. Then, for given β, the solutions for the θ_j^D are obtained by minimizing $H_j^D(\theta_j^D)'\hat{\Delta}_j(\beta)H_j^D(\theta_j^D)$. Since these functions are quadratic in the θ_j^D, closed-form solutions exist. Thus, similarly to the case of strictly exogenous instruments only, we can obtain an optimal GMM estimator of β and θ^* by estimating the two parameter vectors iteratively. The original parameter vector Θ' can then be estimated by $\Theta^D D^{-1}$.

The above iterative procedure can lead to an asymptotically optimal GMM estimator, if we assume the following condition:

$$E(\varepsilon_{is}\varepsilon_{it}\mid x_{S,it},...,x_{S,iT},x_{W,i1},...,x_{W,it},f_i)=E(\varepsilon_{is}\varepsilon_{it}),\ s\geq t,\ t=1,\dots,T,\text{ for all } i. \tag{17}$$

This condition implies that the unconditional variances and covariances of the error terms are the same as their conditional counterparts given all of the current and past values of weakly exogenous regressors, and the time-invariant variables. This condition has been assumed in many empirical studies based on the rational-expectations model (i.e., Hayashi and Sims, 1983; Keane and Runkle, 1992; and Wooldridge, 1996). Studies of dynamic panel data models also often assume this condition strengthening it with the assumption of no autocorrelation in the errors (*i.e.*, Anderson and Hsiao, 1981; Arellano and Bond, 1991; Ahn and Schmidt, 1995; and Blundell and Bond, 1998).

Under (17), we have:

$$E(H_{j,o}^{D}{}'\varepsilon_i\varepsilon_i'H_{k,o}^Dw_{j,i}w_{r,j}')=E(H_{j,o}^{D}{}'\varepsilon_i\varepsilon_i'H_{r,o}^D)E(w_{j,i}w_{r,i}')=H_{j,o}^{D}{}'\Sigma H_{r,o}^DE(w_{j,i}w_{r,i}'),$$

where $H_{j,o}^D=H_j^D(\theta_{j,o}^D)$, and $j,\ r=1,\dots,T\text{-}p$. Now, let D^* be a $(T\text{-}p)\times(T\text{-}p)$ lower triangular matrix such that $D^*D^{*\prime}=[H(\theta_o)'\Sigma H(\theta_o)]^{-1}$ where $\Sigma=E[u_i(\beta_o)u_i(\beta_o)']$. Observe that

$D^{*\prime} H(\theta_o)' \Sigma H(\theta_o) D^* = I_{T-p}$. Thus, when $D = D^*$,

$$Var\left(m_{W,i}^D(\xi_o^D)\right) = diag[E(w_{1,i}w_{1,i}'),...,E(w_{T-p,i}w_{T-p,i}')], \qquad (18)$$

because $H_{j,o}^{D}{}'\Sigma H_{r,o}^{D} = 1$ if $j = r$, and $H_{j,o}^{D}{}'\Sigma H_{r,o}^{D} = 0$ if $j \neq r$. This implies that the weighting matrix $[diag(\hat{\Omega}_{ww,1},...,\hat{\Omega}_{ww,T-p})]^{-1}$ in (16) is optimal for GMM based on (15).

3.3. Identification

In subsection 3.1, we have shown that Assumption (BA.5) is important to identify p_o and β_o. The assumption can be violated if the model (1) contains time-invariant regressors and/or some variables that vary over time but not over individuals (*e.g.*, time-dummy variables). In this subsection, we reexamine these two cases in detail. Bai (2005) consider the same cases when both N and T are large.

Consider the case in which the model (1) contains time-invariant regressors *and* $\Xi_o\alpha_i$ contains a time-invariant component; that is, $y_i = X_i\beta_o + e_T f_i'\gamma_o + (\Xi_o\alpha_i + \varepsilon_i)$, where e_T is a linear combination of the columns of Ξ_o. As discussed above, the parameter vector γ is not identified for this model. The non-identification of the effects of the time-invariant regressors could be a problem when a researcher wishes to estimate the effects of such variables on the dependent variable and test their statistical significance. There is a way to test the insignificance of time-invariant regressors even if $\Xi_o\alpha_i$ contains a time-invariant component. If the variables are important regressors, and if the model (1) is estimated by GMM without those variables, the individual effects $\Xi\alpha_i$ will absorb the effects of the time-invariant regressors, and thus have a time-invariant component. If so, the matrix H, which is orthogonal to Ξ, must be also orthogonal to e_T. Thus, if a Wald test from the GMM estimation of (1) without time-invariant regressors rejects the hypothesis $H(\theta_o)'e_T = 0$, it could be interpreted as evidence that time invariant regressors are not important and all of the individual effects are time-varying.

There is another way to estimate p_o consistently even with time-invariant regressors. We can do so by estimating the model (1) allowing the coefficients of such regressors to vary over time. To be more specific, we rewrite the model with the time-invariant regressors f_i as

$$y_i = X_i\beta + e_T f_i'\gamma + u_i = X_i\beta + \Sigma_{j=1}^{r}(I_T \otimes f_{j,i})e_T\gamma_j + \Xi\alpha_i + \varepsilon_i\,, \tag{19}$$

where $\gamma = (\gamma_1,...,\gamma_{k_f})'$, and k_f is the number of regressors in $f_i = (f_{1,i},...,f_{k_f,i})'$. If we premultiply the model by $H(\theta)'$, we have:

$$H(\theta)'y_i = H(\theta)'X_i\beta + \Sigma_{j=1}^{h_f}(I_{T-p} \otimes f_{j,i})\gamma_j^* + H(\theta)'\varepsilon_i\,, \tag{20}$$

where $\gamma_j^* = H(\theta)'e_T\gamma_j$. When $H(\theta_o)'e_T = 0_{(T-p)\times 1}$, the model (20) cannot identify the original parameter vector γ. However, even for this case, the unrestricted parameter vectors γ_j^* (j = 1, ... , k_f) can be identified. Thus, applying GMM to the model (20) treating γ_j^* (j = 1, ... , k_f) as unrestricted parameter vectors, we can estimate all parameter vectors and the true value of p consistently even if $H(\theta_o)'e_T = 0_{(T-p)\times 1}$. Using the unrestricted GMM regression results, we can also test the hypothesis that $H(\theta_o)'e_T = 0_{(T-p)\times 1}$. If this hypothesis is rejected, it would imply that the individual-effects $\Xi\alpha_i$ do not contain any time-invariant component. Thus, we can estimate the original coefficients γ_j's consistently by reestimating the model (20) with the restrictions $\gamma_j^* = H(\theta)'e_T\gamma_j$ at the value of p estimated from the unrestricted GMM.

However, for two reasons, the unrestricted GMM method is not a complete solution for the non-identification problem. First, too many parameters are involved in the model (20), especially when the number (k_f) of time-invariant regressors is large. This is so because the number of the unrestricted parameters to be estimated is k + (T-p)(p+k_f), where k is the number of time-varying regressors. Thus, the unrestricted GMM estimates of the model (20) may have poor finite-sample properties. Second, if the hypothesis $H(\theta_o)'e_T = 0_{(T-p)\times 1}$ is not rejected, we are still unable to estimate the parameter vector γ consistently.

We now consider the case in which the model (1) contains some variables that vary over time but not over individuals. Denote these variables by G, where the t[th] row of G is g_t'. Using these variables, we can rewrite the model (1) as $y_i = X_i\beta_o + G\lambda_o + (\Xi_o\alpha_i + \varepsilon_i)$. The parameter vector λ can be identified if the number of columns of G (k_g) is smaller than (T-p_o). For this case, the matrix (G, Ξ_o) will be generally of full column rank so that (BA.5) holds. In contrast, if k_g > (T-p_o), λ is not identified. One example of this case is the case with T time-dummy variables (e.g., $G = I_T$). For such cases, the number of columns of the matrix (G, Ξ_o), k_g+p, is greater than T.[7] But since $rank[(G,\Xi_o)] \leq T$, some of the columns of (G, Ξ_o) should be linearly dependent. Let $G = (G_1, G_2)$, where G_1 and G_2 have (k_g+p_o-T) and (T-p_o) columns, respectively. Assume that (G_2, Ξ_o) is full column rank. Then, for some matrices C and D, $G_1 = G_2 C + \Xi_o D$. Then, with $\lambda = (\lambda_1', \lambda_2')'$,

$$G\lambda_o + \Xi_o\alpha_i = G_1(\lambda_{1,o} + d) + G_2(\lambda_{2,o} - Cd) + \Xi_o(\alpha_i - Dd) \equiv G\lambda^* + \Xi_o\alpha_i^*,$$

for any conformable vector d. In this case, for any conformable vector λ, there always exists a matrix H that is orthogonal to both Ξ_o and G.

Differently from the case in which X_i contains time-invariant regressors, we do not have an immediate solution for the non-identification problem of the G-type regressors. If p_o is known, researchers could use up to (T-p_o) G-type regressors in estimation. However, when p_o is unknown, we are unable to determine how many G-type regressors can be used in estimation. The GMM estimation with too many G-type regressors would result in biased estimates of p_o and the regression coefficients. It would be desirable to estimate the model (1) without the

---

[7] The time effects could be identified if we can assume $E(\alpha_i) = 0$. Under this assumption, we can have T additional moment conditions $E[u_i(\beta_o)] = \lambda_o + \Xi_o E(\alpha_i) = \lambda_o$. These moment conditions, with the conditions (5) or (14), are sufficient to identify the vector of time effects λ_o.

G-type regressors leaving their effects to be absorbed by $\Xi\alpha_i$.

4. Monte Carlo Simulations

Our simulations are based on the model with two factors:

$$y_{it} = \beta_{1,o}x_{1,it} + \beta_{2,o}x_{2,it} + [2(\xi_{1t}\alpha_{1i} + \xi_{2t}\alpha_{2i}) + \varepsilon_{it}]. \tag{21}$$

We set $\beta_{1,o} = \beta_{2,o} = 1$. The two time-varying factors ξ_{1t} and ξ_{2t} are drawn from a uniform distribution with range between zero and two [U(0,2)], and the two unobservable individual effects, α_{1i} and α_{2i}, and the error terms ε_{it} are drawn from N (0,1).[8] The two regressors $x_{1,it}$ and $x_{2,it}$ are generated by the following process:

$$x_{j,it} = \rho_{x\alpha}\xi_{jt}\alpha_{ji} + g_{ji} + v^x_{j,it},$$

where $0 \le \rho_{x\alpha} \le 1$, j = 1, 2, and the time-invariant components g_{ji} and time-varying components $v^x_{j,it}$ are drawn from N(0,1). Thus both of the regressors $x_{1,it}$ and $x_{2,it}$ are strictly exogenous with respect to the ε_{it}. Two time-invariant variables are additionally generated by $f_{ji} = \rho_{f\alpha}\alpha_{ji} + g_{ji} + v^z_{ji}$, where $0 \le \rho_{f\alpha} \le 1$, and the v^z_{ji} are from N(0.1). In our experiments, these two variables, in addition to the time-varying regressors $x_{1,it}$ and $x_{2,it}$, are used as instruments.

Each of our experiments consists of 5,000 independent replications.[9] The model (3) is estimated by the iterative method introduced in section 3.1. We consider twenty-one different data generating processes varying the values of N, T, $\rho_{x\alpha}$, and $\rho_{f\alpha}$.

---

[8] As (21) indicates, we generated data multiplying the individual effect components $(\xi_{1t}\alpha_{1i} + \xi_{2t}\alpha_{2i})$ by two. The sizes of the biases in the traditional within and OLS estimators will be smaller (greater) if the components are multiplied by smaller (larger) numbers.

[9] The ξ_{jt} are generated only once and they are repeatedly used for each of the 5,000 replications.

We estimate p using both the sequential test (Proposition 2) and criterion function (Proposition 3) methods. For the sequential method, we have tried two types of significance levels: one group depending on both N and T; and the other only on N. As we have discussed in section 3.1, the sequential test method requires us to adjust the significance level with the size of N. We found from simulations that the sequential method could produce better estimates if the significance level is adjusted with the size of T as well as that of N, if the regressors and individual effects are not highly correlated. We have tried a number of different significance level decision rules. But to save space, we report only two cases:

J1: $b_{N,T}^{1} = 2(T/N)$;

J2: $b_{N}^{2} = 10/N$.

The significance level depends on both N and T for the first case, but only on N for the second case. The functions $b_{N,T}^{1}$ and b_{N}^{2} are chosen so that the significance level equals 10% when N = 100 and T = 5.

For the criterion function method, we consider three functional forms:

AIC: $f(N)=1$, $g(p)=2\,[(q-p)(T-p)-k]$.

BIC1: $f(N,T)=(\ln N)/(\ln T)$, $g(p)=0.75\,[(q-p)(T-p)-k]$.

BIC2: $f(N)=\ln(N)/\ln(5)$, $g(p)=0.75\,[(q-p)(T-p)-k]$.

The AIC criterion is from Cragg and Donald (1997). As they point out, the estimated p by AIC may not be consistent because the value of $f(N)$ is fixed at one. However, we consider AIC to compare its performance with that of the two BIC criteria. Notice that the function f in BIC1 depends on both N and T, while that in BIC2 depends only on N. But the two BIC criteria are equivalent if T = 5. In unreported simulations, we have tried many other functional forms for $f(N)$ and $g(p)$. But their performances were similar to those of BIC1 and BIC2, especially when N is large.

To quantify the influence of the number of instruments on the GMM estimators, we estimate p using three different set of instruments. We index these sets by A, B, and C:

A: $w_{A,i} = (x_{1,i1},...,x_{1,iT}, x_{2,i1},...,x_{2,iT}, f_{1,i}, f_{2,i})'$

B: $w_{B,i} = (\overline{x}_{1,i}, \overline{x}_{2,i}, x_{1,i1}, x_{1,iT}, x_{2,i1}, x_{2,iT}, f_{1,i}, f_{2,i})'$

C: $w_{C,i} = (\overline{x}_{1,i}, \overline{x}_{2,i}, x_{1,i1}, x_{1,iT}, x_{2,i1}, x_{2,iT})'$

Set A contains $(2T+2)$ instruments: $f_{1,i}$, $f_{2,i}$, and all of the leads, current and past values of $x_{1,it}$ and $x_{2,it}$. Thus, for the cases with set A, the number of instruments increases with T. In contrast, the numbers of instruments in B and C do not depend on T. Both the instrument sets B and C contain the individual means, and the first-period and last-period observations of the regressors: that is, six instruments. But set B contains two additional instruments, $f_{1,i}$ and $f_{2,i}$.

Set A contains all the instruments in set B, while all the instruments in C are in B. Thus the GMM estimators of β_1 and β_2 based on sets A and C should be asymptotically most and least efficient, respectively, if they are computed with the true value of p. However, we are unable to predict which set of instruments would produce the most reliable estimate of p.

Table 1 reports the means and root mean squared errors (RMSE) of the five GMM estimators of p: AIC, BIC1, BIC2, J1 and J2. The simulation data for this table are generated with $\rho_{x\alpha} = \rho_{f\alpha} = 1$. The instruments used are those in set A. We consider nine different combinations of N and T: $N = 100$, 300 and 500, and $T = 5$, 7, and 10. For the criterion function methods, we set the maximum value of p at four.

When N = 100, all methods tend to overestimate p. Overall, the BIC2 estimator performs the best in terms of both bias and RMSE. In fact, the performance of the BIC2 estimator is amazingly good: it picks the correct value of p over 95% of the time, for every choice of N and T (although the table does show this fact). The AIC estimator, followed by J2, performs better than the BIC1 and J1 estimators. The (upward) biases in the estimated p by

BIC1 and J1 tend to increase with T. When $N = 300$ or $N = 500$, all of the methods produce smaller biases and RMSEs. The BIC2 estimator still dominates the others, but to a lesser extent. The degrees of overestimation by BIC1 and J1 still tend to increase with T, but only mildly so. Differently from the cases with $N = 100$, the J2 estimator has smaller bias and variability than the AIC estimator. The latter estimator has somewhat larger bias and RMSE than the BIC1 and J1 estimators. Nonetheless, the AIC estimator performs reasonably well regardless of T and N, despite the lack of any theory predicting its consistency. This result is consistent with what Cragg and Donald found from their simulations.

Table 2 reports the means of the relative biases in the estimated β_1 and β_2 obtained using the values of p estimated by the five selection methods. The reported numbers are *relative* biases: that is, $100\times(\hat{\beta}_j - \beta_{j,o})/\beta_{j,o}$, $j = 1,2$. But the true value of β is one, so the relative biases are just bias times 100. The RMSEs of the estimated coefficients are reported in parentheses. For comparison, the table also reports the results from the ordinary least squares (OLS) and the conventional within estimators. Not surprisingly, the OLS and within estimators are biased regardless of sample size. The relative biases in these two estimators are greater than 17 percent in all the cases reported in the table, and greater than 50 percent for most of the cases. In contrast, the GMM estimators have only small biases, even for the cases with $N = 100$. The relative biases in the GMM estimators are smaller than 1% in any case reported. The RMSEs of the GMM estimators are also considerably smaller than those of the OLS and within estimators. The relative bias and variability in the estimated coefficients by BIC1 and J1 are quite comparable to those for the other estimates, even if N is small, despite the fact that the two methods tend to overestimate the true number of factors (Table 1). It appears that the GMM estimates of β_1 and β_2 computed with overestimated p are as reliable as those by the GMM estimates computed with the true number of factors.

The finite sample properties of the GMM estimators of p should depend on the degree of correlation between the individual effects and the instruments (the two regressors, $x_{1,it}$ and $x_{2,it}$, and the two instruments, $f_{1,i}$ and $f_{2,i}$, that are not regressors, but correlated with the effects). Table 3 reports the finite-sample performances of the GMM estimators of *p* for three cases with different levels of correlation between the effects and the regressors (*x*), and between the effects and the instruments that are not regressors (*f*).

For the cases with *N* = 300 and 500, the biases and RMSEs of the five GMM estimators are generally small. They are also not sensitive to the level of correlation between the instruments and the effects. The BIC2 estimator has the smallest bias and RMSE, but only by a small margin. All of the estimators select the correct value of *p* with high probability, but there is a slight tendency to overestimate *p*.

For *N* = 100, the results are rather different. The performances of the five GMM estimators are still not significantly related to the level of correlation (controlled by $\rho_{f\alpha}$) between the effects and the instruments that are not regressors. However, the level of correlation (controlled by $\rho_{x\alpha}$) between the effects and the regressors now makes a difference. These differences are substantial for all of the estimators except BIC1. When the correlation is high we have a slight tendency to overestimate *p*, whereas when the correlation is low we have a clear tendency to underestimate *p*. For example, when $\rho_{x\alpha}$ = 0.2, the BIC2 estimator picks the true value of *p* slightly more than half of the time.[10]

The results reported in Table 3 provide us some limited guidance for empirical studies. If the number of cross-section units is large, all of the five GMM estimators of *p* perform well, although the BIC2 estimator dominates the others by a small margin. In contrast, for data with small *N*, the BIC2 estimator tends to underestimate the number of unobservable factors when the regressors are not highly correlated with the unobservable individual effects. In practice,

[10] Specifically, the estimates of *p* are exactly equal to two (p_o) 57.9% of the time, greater than two 0.6% and smaller than two 41.5%.

we are unable to estimate the correlation between the regressors and the effects, because the latter are not observable. Thus, with small N, researchers may need to estimate the number of factors using different criteria (e.g., BIC1 and BIC2, or J1 and J2) as a robustness check.[11]

The results reported in Tables 1-3 are obtained from the simulations using instrument set A. In order to investigate how the number of instruments could influence the finite-sample properties of the GMM estimators, we also estimate the number of factors and regression coefficients using instrument sets B and C. The results are reported in Tables 4 and 5. Table 4 shows that the GMM estimators computed using a smaller set of instruments tend to overestimate p, especially when N = 100, but only by a small margin. It appears that the estimators of p are not overly sensitive to the number of instruments.

Panel I of Table 5 reports the finite-sample (relative) biases and RMSEs of the GMM estimators of the regression coefficients. The estimators are computed using the estimated numbers of factors by the five decision methods. There is very little bias to discuss.[12] For any estimator, RMSE generally decreases as N increases. In addition, the RMSE generally increases as a smaller number of instruments are used, even if N is large. It appears that the GMM estimators of the regression coefficients have greater variability when they are computed with a smaller number of instruments. For many cases reported in Panel I, the BIC2 estimator has the smallest variability. But as N increases, the RMSEs of the other GMM estimators become closer to those of the BIC2 estimator.

We also investigate the finite-sample size properties of the t-tests based on the coefficient estimates. Panel II of Table 5 shows the rejection rates (sizes) of the t-tests for the hypotheses of $\beta_{1,o} = 1$ or $\beta_{2,o} = 1$. That is, we report the percentages of the 5,000 replications

[11] Of course, all of these estimators are valid *asymptotically*, as N increases with T fixed, and so there is really no expectation that any of these procedures should be reliable when N is small.

[12] The largest relative bias is 0.410, and the true value of the parameter is one, so this corresponds to a bias of only 0.004. With 5000 replications this bias *is* significantly different from zero, but it is numerically unimportant.

in which the true null hypotheses are rejected. Ideally, they should be close to 5% since we test the hypotheses at a 5% nominal significance level. Whatever GMM estimator is used, the t-tests tend to over-reject the correct hypotheses, especially when N = 100. The degree of over-rejection generally (although not always) falls as N increases. When N = 100, the actual sizes of the t-tests computed with the instruments of set C are closest to nominal size (5%). When we use less instruments, the estimators have larger variances (Panel I) but in contrast the asymptotic distribution on which the t-test relies is more reliable and so we are closer to nominal size (Panel II). For the cases with N = 300 or 500, the number of instruments used does not seem to have any noticeable effect on the size of the test.

Our simulation results can be summarized as follows. First, when the sample size is large (N = 300 or more in our experiments), our GMM method estimates the number of factors and regression coefficients quite accurately. The distributions of the estimators are not sensitive to the number of instruments used or to the correlations between the unobservable individual effects and regressors. Second, regardless of the sample size, the GMM estimators of the regression coefficients computed with estimated numbers of factors have much smaller biases and RMSEs than the OLS or the within estimators. Third, when the sample size is small (N = 100 in our experiments), the finite-sample properties of our GMM estimator depend on the criterion functions or the significance levels used to estimate the number of factors. The BIC2 and J2 methods provide more reliable estimates when the regressors are not weakly correlated with the individual effects. But if the regressors are only weakly correlated with individual effects, the BIC1 or J1 methods would be better alternatives. Since the individual effects are unobservable, in practice, we are unable to estimate the correlation between the effects and the regressors. Thus, when the sample size is small, it would be good practice to estimate the number of factors by several different selection methods as a robustness check. Fourth, the number of instruments used does not have substantial effect on the biases and RMSEs of the GMM estimates of the number of factors. When N is small, the coefficient estimates may have larger variances if a smaller number of instrumental variables are used. In contrast, the t-tests applied to estimated coefficients would be more reliable with a smaller number of instruments. The t-tests tend to over-reject correct null hypotheses, but the degree of over-rejection decreases with the sample size.

5. Concluding Remarks

This paper has considered a panel model with time-varying unobservable effects. We consider the case of data with small T and large N. The model resembles the structural equation model which is popularly used in the literature of factor analysis. But the two models are different in two respects. First, the model we consider has a property of the fixed-effects model in that the effects are correlated with the regressors. Second, we do not impose a diagonal structure on the variance matrix of the error terms. Thus, unless we make some parametric restrictions on the correlations between the effects and the regressors, we are unable to use the Jöreskog-type (1967) maximum likelihood method to estimate the model. The GMM method we propose can estimate regression coefficients consistently without imposing such restrictions. The number of factors can be consistently estimated by sequential model testing or BIC-type criterion function methods.

When both N and T are large, the model can be consistently estimated by nonlinear least squares (Bai, 2003, 2005; and Bai and Ng, 2002). However, when the number of time-series observations is small, the nonlinear least squares estimator is inconsistent when the error terms are not white noise. The GMM method we propose avoids this problem. Our simulation results indicate that the proposed GMM estimators generally have good finite-sample properties unless N is relatively small (e.g. $N = 100$).

The model we consider in this paper could be used in a broad range of empirical research. In addition to the examples discussed in the introduction section, the model can be used to estimate the stochastic production frontier model with time-varying technical inefficiencies. In a companion paper (Ahn, Lee and Schmidt, 2005), we use the model and the proposed GMM method to estimate the frontier production function of Indonesian farms. Interested readers may refer to that paper.

References

Ahn, S.C., and P. Schmidt, 1995, Efficient estimation of models for dynamic panel data, *Journal of Econometrics* 68, 5- 27.

Ahn, S.C., Y.H. Lee, and P. Schmidt, 2001, GMM estimation of linear panel data models with time-varying individual effects, *Journal of Econometrics* 101, 219-255.

Ahn, S.C., Y.H. Lee, and P. Schmidt, 2005, Stochastic frontier models with multiple time-varying individual effects, forthcoming in *Journal of Productivity Analysis*.

Altug, S., and R.A. Miller, 1990, Household choices in equilibrium, *Econometrica* 58, 543-570.

Amemiya, T., and T.E. MaCurdy, 1986, Instrumental variable estimation of an error components model, *Econometrica* 54, 869-881.

Anderson, T.W., and C. Hsiao, 1981, Estimation of dynamic models with error components, *Journal of the American Statistical Association* 76, 598-606.

Andrews, D.W.K., 2003, Cross-section regression with common shocks, mimeo, Yale University.

Arellano, M., and S. Bond, 1991, Some tests of specification for panel data: Monte Carlo evidence and application to employment equation, *Review of Economic Studies* 58, 277-297.

Bai, J., 2003, Inferential theory for factor models of large dimensions, *Econometrica* 71, 135-171.

Bai, J., and S. Ng, 2002, Determining the number of factors in approximate factor models, *Econometrica* 70, 191-221.

Bai, J., 2005, Panel data models with interactive fixed effects, mimeo, New York University.

Bauer, P., B.M. Pötscher, and P. Hackl, 1988, Model selection by multiple tests, *Statistics* 19, 39-44.

Blundell, R., and S. Bond, 1998, Initial conditions and moment restrictions in dynamic panel data models, *Journal of Econometrics* 87, 115-143.

Campbell, J.Y., 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373-399.

Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay, 1997, *The Econometrics of Financial Markets*, (Princeton University Press, Princeton, New Jersey).

Caselli, F, G. Esquival and F. Lefort, 1996, Reopening the convergence debate: a new look at cross-country growth empirics, *Journal of Economic Growth* 1, 363-389.

Chamberlain, G., 1992, Efficiency bounds for semiparametric regression, *Econometrica* 60, 567-596.

Chang, Y., 2002, Nonlinear IV unit root tests in panels with cross-section dependency, *Journal of Econometrics* 110, 261-292.

Conley, T.G., 1999, GMM estimation with cross sectional dependence, *Journal of Econometrics* 92, 1-45.

Cornwell, C., and P. Rupert, 1988, Efficient estimation with panel data: an empirical comparison of instrumental variables estimators, *Journal of Applied Econometrics* 3, 149-155.

Cragg, J.C., and S.G. Donald, 1997, Inferring the rank of a matrix, *Journal of Econometrics* 76, 223-250.

Ferson, W.E., and S.R. Foerster, 1994, Finite sample properties of models of moments in latent variable test of asset pricing models, *Journal of Financial Economics* 36, 29-55.

Hansen, P.L., 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029-1054.

Hansen, L.P., J.C. Heaton, and A. Yaron, 1996, Finite sample properties of some alternative GMM estimators, *Journal of Business and Economic Statistics* 14, 262-280.

Hausman, J.A., and W.E. Taylor, 1981, Panel data and unobservable individual effects, *Econometrica* 49, 1377-1398.

Hayashi, H., and C. Sims, 1983, Nearly efficient estimation of time series models with predetermined, but not exogenous, instruments, *Econometrica* 51, 783-798.

Holtz-Eakin, D., W.K. Newey, and H. Rosen, 1988, Estimating Vector Autoregressions with panel data, *Econometrica* 56, 1371-1395.

Islam, N., 1995, Growth empirics: a panel data approach, *Quarterly Journal of Economics* 110, 1127-1170.

Jagannathan, R., and Z. Wang, 1996, The conditional CAPM and the cross-section of expected return, *Journal of Finance* 51, 3 - 53.

Jöreskog, K.G., 1967, Some contributions to maximum likelihood factor analysis, *Psychometrika* 34, 183-202.

Keane, M., and D. Runkle, 1992, On the estimation of panel-data models with serial correlation when instruments are not strictly exogenous, *Journal of Business & Economic Statistics* 10, 1-9.

Kelejian, H.H., and I.R. Prucha, 1999, A generalized moments estimator for autoregressive parameter in a spatial model, *International Economic Review* 40, 509-533.

Kneip, A., R.C. Sickles, and W. Song, 2005, A new panel data treatment for heterogeneity in time trends, Rice University, mimeo.

Lee, Y.H., 1991, Panel data models with multiplicative individual and time effects: Application to compensation and frontier production functions, unpublished Ph.D. dissertation, Michigan State University.

Lee, Y.H., and P. Schmidt, 1993, A production frontier model with flexible temporal variation in technical inefficiency, In: Fried, H., Lovell, C.A.K., Schmidt, S. (Eds.) *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, Oxford.

Magnus, J.R., and H. Neudecker, 1999, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Rev. ed., (John Wiley & Sons, New York).

Mankiw, N.G., D. Romer and D. Weil, 1992, A contribution to the empirics of economic growth, *Quarterly Journal of Economics* 107, 407-437.

Mundlak, Y., 1978, On the pooling of time series and cross section data, *Econometrica* 46, 69-85.

Neyman, J., and E.L. Scott, 1948, Consistent estimates based on partially consistent observations, *Econometrica* 16, 1-32.

Pischke, J-S., 1995, Individual income, incomplete information, and aggregate consumption, *Econometrica* 63, 805-840.

Wooldridge, J.M., 1996, Estimating system of equations with different instruments for different equations, *Journal of Econometrics* 74, 387-405.

Zhou, G., 1994, Analytical GMM tests: asset pricing with time-varying risk premiums, *Review of Financial Studies* 7, 687-709.

Table 1: Monte Carlo Results for Estimation of Number of Factors

The simulation data are generated with $\rho_{x\alpha} = \rho_{f\alpha} = 1.0$ and two factors ($p_o = 2$). The instruments used are two time-invariant variables and all of the leads and lags of the two time-varying regressors (set A).

		AIC	BIC1	BIC2	J1	J2
N	***T***	*p*	*p*	*p*	*p*	*p*
100	5	2.063[a] (0.253)[b]	2.033 (0.190)	2.033 (0.190)	2.090 (0.321)	2.090 (0.321)
	7	2.040 (0.202)	2.157 (0.412)	2.015 (0.122)	2.148 (0.417)	2.100 (0.341)
	10	2.008 (0.089)	2.636 (0.930)	2.001 (0.032)	2.240 (0.555)	2.070 (0.367)
300	5	2.084 (0.297)	2.007 (0.084)	2.007 (0.084)	2.033 (0.187)	2.033 (0.187)
	7	2.069 (0.266)	2.025 (0.158)	2.001 (0.032)	2.040 (0.210)	2.027 (0.167)
	10	2.053 (0.232)	2.134 (0.379)	2.000 (0.000)	2.057 (0.255)	2.027 (0.173)
500	5	2.085 (0.297)	2.003 (0.055)	2.003 (0.055)	2.020 (0.145)	2.020 (0.145)
	7	2.079 (0.283)	2.009 (0.095)	2.001 (0.032)	2.020 (0.141)	2.013 (0.114)
	10	2.072 (0.272)	2.062 (0.253)	2.000 (0.000)	2.041 (0.212)	2.020 (0.152)

[a] Average of estimates.
[b] The numbers in (•) are root mean squared errors.

Table 2: Monte Carlo Results for Estimation of Regression Coefficients

The simulation data are generated with $\rho_{x\alpha} = \rho_{f\alpha} = 1.0$ and two factors ($p_o = 2$). The instruments used are two time-invariant variables and all of the leads and lags of the two time-varying regressors (set A).

		OLS		Within		AIC		BIC1		BIC2		J1		J2	
N	***T***	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
100	5	53.22[a]	74.27	58.22	17.34	0.071	0.084	0.069	0.080	0.069	0.080	0.072	0.062	0.072	0.062
		(0.542)[b]	(0.750)	(0.588)	(0.190)	(0.060)	(0.061)	(0.059)	(0.060)	(0.059)	(0.060)	(0.061)	(0.062)	(0.061)	(0.062)
	7	65.66	79.69	51.27	54.82	-0.031	-0.040	-0.005	-0.028	-0.028	-0.061	-0.003	-0.011	-0.008	-0.035
		(0.665)	(0.803)	(0.518)	(0.554)	(0.043)	(0.044)	(0.044)	(0.046)	(0.042)	(0.043)	(0.044)	(0.045)	(0.043)	(0.045)
	10	81.59	100.37	56.77	48.96	-0.129	-0.215	-0.023	-0.134	-0.131	-0.221	-0.050	-0.160	-0.032	-0.121
		(0.823)	(1.009)	(0.572)	(0.495)	(0.036)	(0.036)	(0.040)	(0.039)	(0.036)	(0.036)	(0.038)	(0.037)	(0.037)	(0.036)
300	5	90.54	51.72	35.20	61.57	0.007	-0.014	0.028	0.000	0.028	0.000	0.021	0.004	0.021	0.004
		(0.907)	(0.521)	(0.355)	(0.618)	(0.034)	(0.034)	(0.032)	(0.032)	(0.032)	(0.032)	(0.033)	(0.032)	(0.033)	(0.032)
	7	78.44	57.96	62.85	66.52	0.030	-0.016	0.036	-0.015	0.029	-0.023	0.041	-0.018	0.035	-0.016
		(0.786)	(0.583)	(0.630)	(0.667)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
	10	99.15	79.55	32.14	49.58	-0.073	-0.053	-0.072	-0.047	-0.074	-0.058	-0.072	-0.055	-0.073	-0.055
		(0.993)	(0.798)	(0.323)	(0.497)	(0.020)	(0.020)	(0.021)	(0.021)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
500	5	93.05	81.11	35.68	70.07	0.050	-0.048	0.040	-0.039	0.040	-0.039	0.050	-0.043	0.050	-0.043
		(0.932)	(0.813)	(0.359)	(0.702)	(0.026)	(0.026)	(0.024)	(0.025)	(0.024)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
	7	66.17	59.46	47.13	37.66	-0.009	0.043	-0.017	0.041	-0.015	0.042	-0.013	0.048	-0.013	0.046
		(0.663)	(0.596)	(0.472)	(0.378)	(0.020)	(0.020)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
	10	88.86	73.80	44.69	56.35	-0.028	-0.020	-0.025	-0.021	-0.033	-0.024	-0.032	-0.021	-0.034	-0.022
		(0.890)	(0.740)	(0.448)	(0.564)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)

[a] Average of relative biases in estimates (%).
[b] The numbers in (•) are root mean squared errors.

Table 3: Monte Carlo Results for Estimation of Number of Factors with Different Correlations between Individual Effects and Instruments

The simulation data are generated with $T = 7$ and two factors ($p_o = 2$). The instruments used are two time invariant variables, all of the leads and lags of the two time-varying regressors (set A).

		AIC	BIC1	BIC2	J1	J2
		p	p	p	p	p
N	$\rho_{f\alpha} = 1.0$					
100	$\rho_{x\alpha} = 1.0$	2.040[a]	2.157	2.015	2.148	2.100
		(0.202)[b]	(0.412)	(0.122)	(0.417)	(0.341)
	$\rho_{x\alpha} = 0.5$	1.877	2.074	1.712	1.920	1.807
		(0.418)	(0.396)	(0.570)	(0.552)	(0.586)
	$\rho_{x\alpha} = 0.2$	1.790	2.029	1.555	1.850	1.732
		(0.503)	(0.417)	(0.729)	(0.602)	(0.635)
300	$\rho_{x\alpha} = 1.0$	2.069	2.025	2.001	2.040	2.027
		(0.266)	(0.158)	(0.032)	(0.210)	(0.167)
	$\rho_{x\alpha} = 0.5$	2.069	2.026	2.001	2.038	2.025
		(0.268)	(0.161)	(0.032)	(0.197)	(0.161)
	$\rho_{x\alpha} = 0.2$	2.066	2.023	2.001	2.038	2.026
		(0.265)	(0.155)	(0.032)	(0.200)	(0.164)
500	$\rho_{x\alpha} = 1.0$	2.079	2.009	2.001	2.020	2.013
		(0.283)	(0.095)	(0.032)	(0.141)	(0.114)
	$\rho_{x\alpha} = 0.5$	2.076	2.010	2.001	2.026	2.018
		(0.281)	(0.100)	(0.032)	(0.164)	(0.138)
	$\rho_{x\alpha} = 0.2$	2.077	2.011	2.001	2.026	2.017
		(0.281)	(0.105)	(0.032)	(0.167)	(0.134)
N	$\rho_{x\alpha} = 1.0$					
100	$\rho_{f\alpha} = 1.0$	2.040	2.157	2.015	2.148	2.100
		(0.202)	(0.412)	(0.122)	(0.417)	(0.341)
	$\rho_{f\alpha} = 0.5$	2.043	2.155	2.017	2.147	2.093
		(0.207)	(0.407)	(0.130)	(0.418)	(0.332)
	$\rho_{f\alpha} = 0.2$	2.044	2.159	2.017	2.150	2.100
		(0.212)	(0.412)	(0.134)	(0.424)	(0.342)
300	$\rho_{f\alpha} = 1.0$	2.069	2.025	2.001	2.040	2.027
		(0.266)	(0.158)	(0.032)	(0.210)	(0.167)
	$\rho_{f\alpha} = 0.5$	2.074	2.025	2.001	2.043	2.030
		(0.279)	(0.161)	(0.032)	(0.212)	(0.176)
	$\rho_{f\alpha} = 0.2$	2.073	2.025	2.001	2.043	2.030
		(0.274)	(0.158)	(0.032)	(0.214)	(0.176)
500	$\rho_{f\alpha} = 1.0$	2.079	2.009	2.001	2.020	2.013
		(0.283)	(0.095)	(0.032)	(0.141)	(0.114)
	$\rho_{f\alpha} = 0.5$	2.085	2.011	2.001	2.027	2.021
		(0.295)	(0.105)	(0.032)	(0.167)	(0.148)
	$\rho_{f\alpha} = 0.2$	2.083	2.012	2.001	2.028	2.022
		(0.292)	(0.110)	(0.032)	(0.170)	(0.148)

[a] Average of estimates

[b] The numbers in (•) are root mean squared errors

Table 4: Monte Carlo Results for Estimation of Number of Factors with Different Instrument Sets

The simulation data are generated with $\rho_{x\alpha} = \rho_{f\alpha} = 1.0$, $T = 7$, and two factors ($p_o = 2$). Three different sets of instruments are used. Set A contains two time-invariant variables and all of the leads and lags of the two time-varying regressors ((2T+2) instruments). Set B includes eight instruments: two time-invariant instruments, the individual means, and the first- and last-period values of the two time-varying regressors. Finally, set C contains the same instruments as set B, except the two time-invariant instruments: that is, six instruments.

N	Instrument Type	AIC	BIC1	BIC2	J1	J2
		P	p	p	p	p
100	A	2.040[a] (0.202)[b]	2.157 (0.412)	2.015 (0.122)	2.148 (0.417)	2.100 (0.341)
	B	2.128 (0.385)	2.262 (0.562)	2.074 (0.298)	2.121 (0.396)	2.071 (0.333)
	C	2.128 (0.387)	2.234 (0.526)	2.085 (0.321)	2.130 (0.405)	2.079 (0.332)
300	A	2.069 (0.266)	2.025 (0.158)	2.001 (0.032)	2.040 (0.210)	2.027 (0.167)
	B	2.144 (0.400)	2.083 (0.297)	2.015 (0.122)	2.039 (0.202)	2.030 (0.176)
	C	2.141 (0.396)	2.085 (0.300)	2.026 (0.161)	2.045 (0.219)	2.032 (0.182)
500	A	2.079 (0.283)	2.009 (0.095)	2.001 (0.032)	2.020 (0.141)	2.013 (0.114)
	B	2.161 (0.427)	2.048 (0.224)	2.007 (0.084)	2.026 (0.161)	2.019 (0.138)
	C	2.161 (0.424)	2.056 (0.245)	2.011 (0.105)	2.028 (0.170)	2.019 (0.138)

[a] Average of estimates

[b] The numbers in (•) are root mean squared errors

Table 5: Monte Carlo Results for Estimation of Regression Coefficients with Different Instruments

The simulation data are generated with $\rho_{x\alpha} = \rho_{f\alpha} = 1.0$, $T = 7$, and two factors ($p_o = 2$). Three different sets of instruments are used. Set A contains two time-invariant variables and all of the leads and lags of the two time-varying regressors (($2T$+2) instruments). Set B includes eight instruments: two time-invariant instruments, the individual means, and the first- and last-period values of the two time-varying regressors. Finally, set C contains the same instruments as set B, except the two time-invariant instruments: that is, six instruments.

Panel I: Bias and RMSE

		AIC		BIC1		BIC2		J1		J2	
N	Inst. Type	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
100	A	-0.031[a]	-0.040	-0.005	-0.028	-0.028	-0.061	-0.003	-0.011	-0.008	-0.035
		(0.043)[b]	(0.044)	(0.044)	(0.046)	(0.042)	(0.043)	(0.044)	(0.045)	(0.043)	(0.045)
	B	0.410	0.206	0.409	0.275	0.380	0.139	0.407	0.089	0.424	0.031
		(0.101)	(0.098)	(0.120)	(0.123)	(0.097)	(0.089)	(0.098)	(0.088)	(0.094)	(0.079)
	C	0.247	-0.276	0.463	-0.567	0.225	-0.119	0.173	-0.277	0.165	-0.086
		(0.142)	(0.136)	(0.175)	(0.167)	(0.124)	(0.118)	(0.133)	(0.123)	(0.114)	(0.108)
300	A	0.030	-0.016	0.036	-0.015	0.029	-0.023	0.041	-0.018	0.035	-0.016
		(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
	B	-0.111	-0.257	-0.096	-0.182	-0.080	-0.154	-0.128	-0.174	-0.129	-0.167
		(0.055)	(0.050)	(0.052)	(0.041)	(0.047)	(0.038)	(0.048)	(0.040)	(0.048)	(0.039)
	C	-0.430	-0.418	-0.255	-0.225	-0.168	-0.149	-0.215	-0.158	-0.179	-0.156
		(0.075)	(0.062)	(0.062)	(0.047)	(0.055)	(0.039)	(0.058)	(0.040)	(0.056)	(0.040)
500	A	-0.009	0.043	-0.017	0.041	-0.015	0.042	-0.013	0.048	-0.013	0.046
		(0.020)	(0.020)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
	B	-0.083	-0.032	0.071	-0.031	0.041	0.002	0.043	0.012	0.056	0.002
		(0.063)	(0.050)	(0.038)	(0.035)	(0.031)	(0.031)	(0.031)	(0.032)	(0.031)	(0.031)
	C	-0.267	-0.448	0.032	-0.060	0.046	-0.003	0.042	-0.024	0.047	-0.015
		(0.078)	(0.084)	(0.045)	(0.043)	(0.031)	(0.031)	(0.032)	(0.033)	(0.031)	(0.032)

Panel II: Size of t-Statistics

		AIC		BIC1		BIC2		J1		J2	
N	Inst. Type	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	B_1	β_2
100	A	6.52[c]	6.98	6.96	7.32	6.34	6.92	7.04	7.12	6.80	7.12
	B	6.42	6.68	7.74	7.66	6.34	6.22	6.70	6.46	6.56	6.24
	C	6.10	5.96	6.94	6.42	5.66	5.70	6.10	5.82	5.62	5.62
300	A	5.56	5.76	5.28	5.44	5.12	5.30	5.46	5.54	5.36	5.44
	B	6.70	6.04	6.34	5.82	5.78	5.32	6.00	5.72	5.90	5.54
	C	6.84	5.98	6.38	5.74	5.92	5.30	6.04	5.44	5.96	5.38
500	A	5.76	5.88	5.56	5.44	5.48	5.44	5.56	5.48	5.50	5.48
	B	6.60	6.32	5.56	5.82	5.18	5.56	5.38	5.68	5.36	5.64
	C	6.08	6.24	5.76	5.98	5.26	5.54	5.40	5.68	5.38	5.54

[a] Average of relative biases in estimates (%).

[b] The numbers in (•) are root mean squared errors.

[c] Rejection rate of H_o: $\beta_1 = 1$ at 5% significance level.